

Camera Calibration from Dynamic Silhouettes Using Motion Barcodes

Gil Ben-Artzi Yoni Kasten Shmuel Peleg Michael Werman
 School of Computer Science and Engineering
 The Hebrew University of Jerusalem, Israel

Abstract

Computing the epipolar geometry between cameras with very different viewpoints is often problematic as matching points are hard to find. In these cases, it has been proposed to use information from dynamic objects in the scene for suggesting point and line correspondences.

We propose a speed up of about two orders of magnitude, as well as an increase in robustness and accuracy, to methods computing epipolar geometry from dynamic silhouettes. This improvement is based on a new temporal signature: motion barcode for lines. Motion barcode is a binary temporal sequence for lines, indicating for each frame the existence of at least one foreground pixel on that line. The motion barcodes of two corresponding epipolar lines are very similar, so the search for corresponding epipolar lines can be limited only to lines having similar barcodes. The use of motion barcodes leads to increased speed, accuracy, and robustness in computing the epipolar geometry.

1. Introduction

Calibration of multi-camera systems is normally computed by finding corresponding feature points between images taken by these cameras. When not enough feature points can be found, e.g. when the camera viewpoints vary greatly, the epipolar geometry can be computed from silhouettes of moving objects that are visible in videos captured by the two cameras. The silhouettes at one time instance are used to suggest matching epipolar lines which are used to propose a fundamental matrix that is verified over all frames.

The best methods for computing the fundamental matrix use tangents to the dynamic silhouette as candidates for epipolar lines [28]. Our approach presents a speedup of about two orders of magnitude for these methods, and significantly improves accuracy and robustness. This speedup is obtained by requiring candidates for matching epipolar lines to share a temporal signature (motion barcode).

Motion barcodes were first introduced for points in [5]. The motion barcode of a line is a binary temporal sequence,

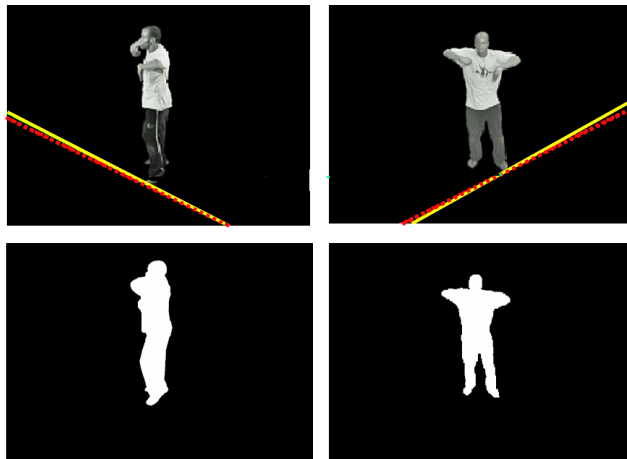


Figure 1. When two cameras have very different viewpoints as in this example, appearance can not be used for calibration. Instead, calibration is possible from matching pairs of epipolar lines that can be extracted efficiently from moving silhouettes. The yellow lines are the epipolar lines proposed by our method, while the red lines are the ground truth epipolar lines. The corresponding silhouettes are displayed at the bottom.

indicating for each frame the existence of at least one foreground pixel on that line. We show that correlation between the motion barcodes of corresponding epipolar lines is high. By testing as possible matches only pairs of lines whose motion barcode correlation is high, a speedup by about two orders of magnitude is obtained. Figure 1 shows matching epipolar lines extracted using our approach. Following [28], we use a RANSAC approach to test possible matching pairs of epipolar lines, and compute the epipolar geometry.

This paper is organized as follows. Section 1.1 describes relevant prior work. Section 2 introduces the theoretical background. Section 3 presents the motion barcodes of lines. Section 4 shows how to match epipolar lines based on dynamic silhouettes. Section 5 presents an iterative computation of the fundamental matrix based on the motion barcode. Section 6 shows our results on both synthetic and real sequences.

1.1. Prior Work

Extracting geometrical information from the motion of silhouettes include shape-from-silhouettes [7, 14, 3] and camera calibration [18, 28, 6, 26, 34]. In shape-from-silhouettes, the goal is to recover the visual hull [20, 23] of the object. If the cameras are calibrated, this task is relatively clear as each individual viewing cone [15] can be backprojected and the visual hull is the intersection of these cones.

The case of uncalibrated cameras has also been investigated, where the goal is to recover epipolar geometry. The first step is to establish correspondences between special points on the silhouettes boundaries, called *frontier points* [9], across the different views. These points are images of object points that are tangent to an epipolar plane. Given corresponding frontier points, spatial constraints resulting from matching epipolar tangents [27] are used to recover the epipolar geometry.

Matching corresponding frontier points and silhouette tangents can be found using robust estimation procedures such as RANSAC [13]. Matching frontier points, or directions of four epipolar tangent lines, are initially guessed. Furukawa et al. [16], assuming orthographic projection, match frontier points using RANSAC. They used the distances between parallel tangent lines on the silhouettes as a geometric measure for matching. Given the epipoles and the accurate tangent envelope of the silhouettes, frontier points can be easily matched using two outermost epipolar tangents. This property was deployed by Wong and Cipolla [31] for turntable motion. The most relevant previous work is Sinha and Pollefeys [28], addressing projective projection. They propose a RANSAC based search of possible epipoles, where a proposed epipole in each of the two corresponding images is generated from the intersection of two lines randomly selected from the tangent envelope.

Calibration without explicitly matching tangent epipolar lines has also been considered. [6] used constraints based on the back projection of silhouettes boundaries in multiple views. [33] jointly optimized the 3D position of frontier points and the camera parameters in a bundle adjustment. However, both methods require a good initialization of silhouette boundaries and camera parameters. Hernandez [18] also proposed constraints based on the back projection of silhouettes for maximizing silhouette coherence, but his method is limited to turntable motion.

Binary temporal signatures of pixels which are based on the motion of the objects in the scene have been previously introduced. Ermis et al. [12] deploy such features to find accurate correspondence between pixels across distributed cameras with the assumption of a distant, almost planar, scene. Drouin et al. [11] matched 2D points between a video projector and a digital camera. They require a planar surface and the same ordering of pixels across views. The

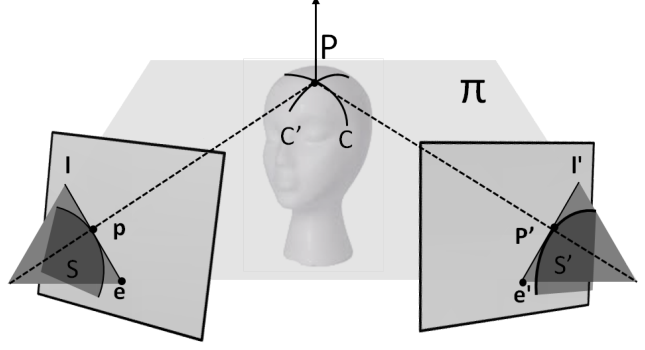


Figure 2. The geometry of two views with a tangent epipolar plane and a frontier point. Π is the tangent epipolar plane and l, l' are tangent epipolar lines in Π . The epipolar plane is tangent to the object at the frontier points P .

closest work related to the motion barcode is [25] where the *line signal*, the sum of intensity values of pixels on epipolar lines after background subtraction, was defined. This was used for video synchronization. The line signal depends on motion and color and was used assuming known calibration. Ben-Artzi et al. [5] introduce a method to match events across different views even in the case of significant parallax and occlusions. However, that approach can not be applied to match pixels for camera calibration, as its localization is very inaccurate, as explained in Section 3. Using a temporal histogram as a temporal feature of a pixel is introduced in [19], but it is effective only for objects that are static for substantial periods.

2. Theoretical Background

The geometric relation between corresponding silhouettes across views is based on frontier points and epipolar tangencies [22, 8, 21].

The geometry of two views containing silhouettes is presented in Fig. 2. For the rest of the paper let *candidate points* be image points that are on the boundary of the silhouette as well as on the boundary of its convex hull. C and C' are the contours of the object in 3D. These contours project to silhouette boundaries S and S' . The two contours intersect in the frontier point P . The projections of P onto the two views are the candidate points, p and p' . The 3 points P, p, p' span a *tangent* epipolar plane Π between the two views. The points p and p' must lie on the corresponding tangent lines l and l' and the point P is the location where the tangent epipolar plane Π is tangent to the surface, e, e' are the epipoles. The frontier points are the only true corresponding points between the boundaries S and S' . If we have accurate tangents to the silhouettes and the location of the epipole is known, then the epipolar tangent lines give the corresponding points p and p' . This idea

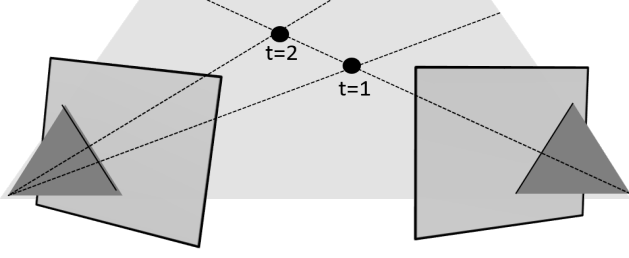


Figure 3. In dynamic scenes, the geometrical relation between pixels is characterized only up to corresponding epipolar lines. Each pixel in one video can correspond to different pixels at different times in the other video. For stationary cameras, the different pixels in the second view will always reside on the epipolar line.

was traditionally used as a *spatial* cost function between corresponding tangential epipolar lines and points. See for example [16, 28].

Finding frontier points without the epipole locations is difficult [24]. For a video sequence, when the location of at least two frontier points are known, their tangent lines are epipolar lines. They can be used to calculate the epipole and the location of the other frontier points in all the other frames. Alternatively, if the epipoles are known we can use the tangent lines to the silhouettes to locate the frontier points. It follows that either the frontier points or the epipoles are needed in order to extract the epipolar lines.

Here we introduce a different approach which does not require prior knowledge of either frontier points or epipoles in order to extract matching epipolar lines. We directly compute epipolar line correspondences using the motion observed simultaneously by them.

3. Motion Barcode: Temporal Signature of Lines

Given two frames captured at the same time from different viewpoints, two corresponding pixels view a single 3D point. However, in a dynamic scene, a pixel in one view is bound to correspond to different pixels of the other view at different times, located on the corresponding epipolar line.

Fig. 3 illustrates a typical case. At time $t = 1$, a single pixel in the right view corresponds to some pixel in the left view. At time $t = 2$, due to the motion of the object in scene, the same pixel corresponds to a different pixel in the other view. For video sequences captured by stationary cameras the corresponding pixels will always reside on corresponding epipolar lines.

It follows that if an epipolar line contains *at least* one silhouette pixel at time t , then its corresponding epipolar line should contain such a pixel at the *same time*. This is illustrated in Fig. 4. For time $t = 2, 3$ there are points from objects that project onto the corresponding epipolar line. If

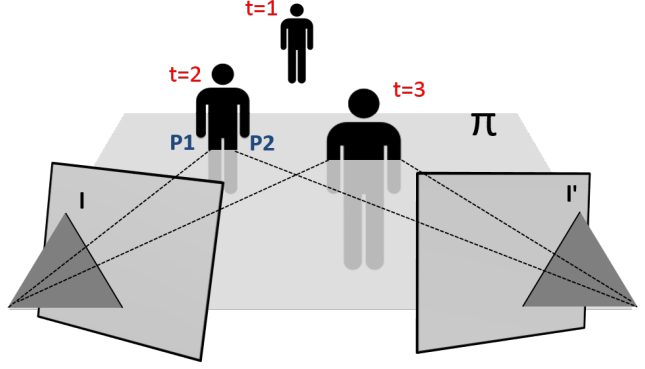


Figure 4. The motion barcodes of two corresponding epipolar lines, l and l' , in a video of a moving person at three time instances. If a point on an epipolar line is a projection of a foreground point at time t , then there exists a point on its corresponding epipolar line which is also a projection of a foreground point at the same time. In the figure, at time $t = 2, 3$ the two corresponding epipolar lines contain a point from a silhouette. This can be a different 3D point due to viewpoint differences, e.g. P_1, P_2 . The motion barcode of both epipolar lines in this figure, b_l and $b_{l'}$, is $[0,1,1]$.

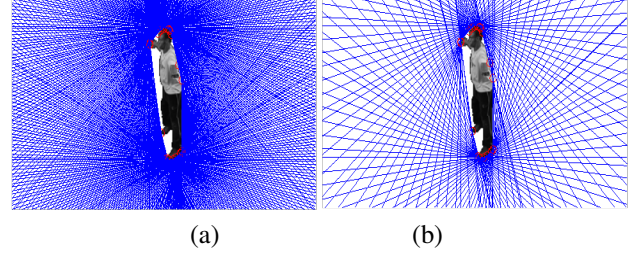


Figure 5. Uniform sampling of lines from the tangent envelope. (a) Lines sampled every 1° , (b) Lines sampled every 4°

a point on line l is part of a silhouette, this point or another silhouette point occluding it, will be seen on line l' . Alternatively, the silhouette could be blocked by a background object or be out of the frame.

The motion barcode of a line l , $b_l(t)$, indicates for each line l in frame t the existence of at least one foreground pixel on that line. $b_l(t) = 1$ if the line intersects a silhouette, and $b_l(t) = 0$ otherwise. The motion barcodes of two corresponding epipolar lines is very similar. Differences occur only in cases of occlusions.

The *temporal* similarity between two lines l and l' is defined as the correlation between their motion barcodes;

$$d_t(l, l') = \text{corr}(b_l(t), b_{l'}(t)) \quad (1)$$

4. Epipolar Geometry by Matching Lines

The epipolar geometry can be computed from 3 pairs of corresponding epipolar lines [17]. The search space for

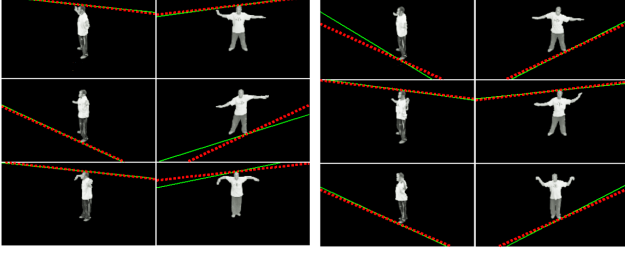


Figure 6. Finding corresponding epipolar lines by their motion barcode. Every pair of frames contributes one possible match. The dashed lines are the true epipolar lines and the green lines are the candidate pairs having highest barcode correlation.

matching epipolar lines across views is very large if we consider all possible pairs of lines. This search can be reduced to fewer lines by using only *candidate lines*, lines tangent to the *tangent envelope* of the silhouette. The tangent envelope includes points that are on the silhouette boundary as well as on the boundary of its convex hull. We follow the work of [28], checking for possible correspondence only lines on the tangent envelope. Using only candidate lines is justified as the projection of the frontier point is on the tangent envelope.

We select several corresponding pairs of frames from the video sequences, so that the pairs will be sufficiently different from each other. For each pair of frames, we sample K candidate lines from the tangent envelope of its silhouettes. We compute the correlation between the motion barcodes for all pairs of candidate lines from the two corresponding images. This results in K^2 correlations per each pair of frames. From every pair of frames we select the single pair of epipolar lines with the highest barcode correlation. Fig. 5 shows all candidate lines, and Fig. 6 shows the pairs of candidate lines having highest barcode correlation. We compute the epipolar geometry from three matching pairs based on [17]. The computation is by RANSAC similarly to [28]. The fundamental matrix is then fully optimized as described in Section 5.

The matching is carried out in two phases. An offline phase where the motion barcodes of the tangent lines are computed. In the online phase, the actual matching is carried out by computing the correlation between motion barcodes of pairs of lines, and computing the epipolar geometry using RANSAC. The overall efficiency depends mainly on accuracy of candidate matches as it effects the number of required iterations in the RANSAC phase. Details are in Section 6.

5. Temporal Optimization of the Fundamental Matrix

Existing optimization techniques for computing the fundamental matrix are based on minimizing a spatial cost function without taking into account the temporal dimension. We present a technique based on both spatial and temporal cost functions (Eq. 1).

We assume a set of corresponding points, presumably the projection of frontier points $\{(x_i, x'_i)\}_{i=1}^M$, a set of corresponding epipolar lines $\{(l_i, l'_i)\}_{i=1}^N$ and an initial estimation of the fundamental matrix F . The optimization is iterative. In the first step we optimize the point correspondences based on the lines, using the geometric reprojection error [17] as the spatial cost function. In the second step we optimize the epipolar line correspondences based on the given points, using the temporal cost function (Eq. 1). We optimize the directions of the epipolar lines for each pair of corresponding points. Based on the lines matched in this step, we estimate epipoles and an epipolar line homography. We then evaluate a set of corresponding points and obtain an estimation of the fundamental matrix. The process is described in the following:

- Step one:

1. Minimize reprojection error based on $\{(x_i, x'_i)\}_{i=1}^M$.

$$\sum_i d(x_i, \hat{x}_i)^2 + d(x'_i, \hat{x}'_i)^2 \quad s.t. \quad \hat{x}_i \hat{F} \hat{x}'_i = 0$$

This minimization is by the Levenberg-Marquardt procedure and gives a new set of points and fundamental matrix.

2. set $l_i = \hat{F} \hat{x}'_i, l'_i = \hat{F}^T \hat{x}_i$.

- Step two:

1. For each pair of lines, minimize

$$C_l(\hat{l}, \hat{l}') = d_s(l, \hat{l}) + d_s(l', \hat{l}') - d_t(\hat{l}, \hat{l}') \quad (2)$$

d_s measures the angular deviation between lines, and d_t is the barcode correlation (Eq. 1). d_s ensures the lines are within an angle difference of no more than Θ . The choice of Θ will be discussed next. \hat{l}, \hat{l}' are sampled uniformly from $[-\Theta, \Theta]$ around l, l' . We take the maximal match and if we have more than one maximum, the one with the minimal angle difference is selected.

2. Estimate new epipoles e, e' and epipolar line homography from $\{\hat{l}_i, \hat{l}'_i\}$.
3. Set $\{x_i, x'_i\}$ by projecting onto the nearest l, l' . Estimate F from epipoles and lines homography.

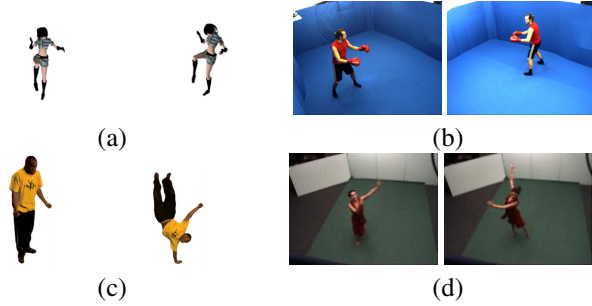


Figure 7. The datasets used in the experiments. (a) The synthetic Kung-Fu girl dataset. (b) The Boxer dataset. (c) The Street Dancer dataset (d) The Dancing Girl dataset.

The process terminates when the deviation of the estimated epipoles is small enough or a maximum number of iterations is exceeded.

The choice of the angular tolerance Θ defines the region where we look for the newly estimated lines. It depends on the epipolar envelope and the required probability for locating the line [30]. Direct modeling of epipolar envelope is difficult and therefore it is empirically evaluated, see [10, 32, 17]. In our implementation we set Θ to 0.2° which results in an accurate estimation. This reflects our assumption that the distortion is low. The specific choice can be adjusted according to the needs.

6. Experiments

Our approach was validated on synthetic and real sequences. We compared our method with the state of the art method [28], where the fundamental matrix is computed by RANSAC-based sampling of epipolar lines. The evaluation was done with the following datasets: the Kung-Fu girl [2], Boxer [4], Street Dancer [29] and Dancing Girl [1]. Fig. 7 shows images from the datasets and Table 1 gives the details.

We compared the accuracy and efficiency of the two methods. The accuracy of the fundamental matrix in all experiments is measured by the symmetric epipolar distance (error) [17] using ground truth matching points. The symmetric epipolar distance is the distance between each point and the epipolar line corresponding to the other point. The acquisition of the ground truth points is discussed in Subsection 6.3.

Dataset	Type	Camera Pairs	Frames
KungFu Girl	Synthetic	300	200
Boxer	Real	6	778
Street Dancer	Real	15	250
Dancing Girl	Real	28	200

Table 1. Dataset properties

No. of Non-Linear Optimizations Needed to Reach a Desired Accuracy							
Sym Epipolar Distance		1.5	1	0.8	0.5	0.4	0.3
Kung-Fu	Ours	1	2	4	23	71	302
	Sinha	19	65	134	822	1989	8659
Street Dancer	Ours	3	7	20	255	616	1233
	Sinha	37	159	340	1871	7485	-
Dancing Girl	Ours	2	4	9	129	918	13776
	Sinha	36	149	388	13972	-	-
Boxer	Ours	2	5	12	111	996	-
	Sinha	333	2994	2994	-	-	-

Table 2. The expected number of non-linear optimizations required to reach a given accuracy of the fundamental matrix. Accuracy is measured using symmetric epipolar distance with respect to ground-truth points. The best hypothesis is selected every 1000 RANSAC iterations, and is further optimized using non-linear (LM) method. In each dataset, the number of optimizations is averaged over all camera pairs. Empty cells indicate that the required accuracy was not attained.

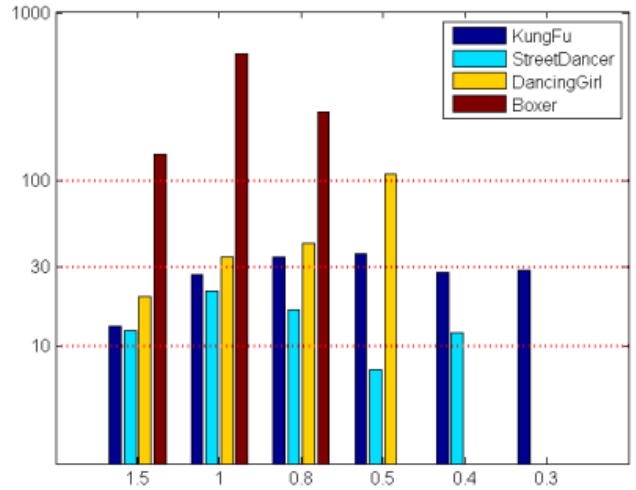


Figure 8. The ratio between our method and Sinha [28] of the number of non-linear optimization procedures required to reach a given fundamental matrix accuracy. The horizontal axis is the accuracy in terms of the desired symmetric epipolar distance of ground truth points.

The efficiency of the methods is evaluated as follows. In both methods the fundamental matrix is computed using RANSAC sampling of epipolar lines. In each iteration, the symmetric epipolar distance of each hypothesis is evaluated. Every 1000 RANSAC iterations the best hypothesis is selected and optimized using the non-linear Levenberg-Marquardt (LM) optimization procedure as in [28]. The efficiency is measured by the number of non-linear optimization procedures required to reach a given accuracy (error). The less non-linear optimization procedures the more efficient the method is. A detailed description is in Subsection 6.3.

There is a difference in the error used during RANSAC and the error we use for final evaluation. During RANSAC,

		Symmetric Epipolar Distance					
RANSAC Hypotheses		1K	2K	5K	10K	20K	100K
Kung-Fu	ours	1.11	0.85	0.64	0.54	0.47	0.35
	Sinha	4.9	3.41	2.12	1.63	1.29	0.78
Street Dancer	ours	1.93	1.29	0.97	0.85	0.75	0.59
	Sinha	4.31	3.35	2.4	1.96	1.59	1.01
Dancing Girl	Ours	1.4	1.09	0.83	0.72	0.63	0.49
	Sinha	6.28	4.57	2.96	2.15	1.6	1
Boxer	Ours	1.63	1.46	0.85	0.74	0.65	0.48
	Sinha	7.06	5.82	4.02	3.37	2.8	1.86

Table 3. Accuracy reached by each method for a fixed number of RANSAC samples. Accuracy is after a non-linear optimization phase, measured with respect to ground-truth points.

	Ours	Sinha
Kung-Fu	0.26	0.51
Street Dancer	0.36	0.62
Dancing Girl	0.41	0.72
Boxer	0.39	1.33

Table 4. The best accuracy reached by each method on all camera pairs in each dataset after 500K RANSAC hypotheses. The accuracy is the median over all camera pairs of the best symmetric epipolar distance reached after the non-linear optimization phase.

the quality of an hypothesis is evaluated based on inliers, as ground truth is unknown. This error is usually lower from the error of ground truth points. We used ground-truth points for a non biased evaluation.

Efficiency The expected number of non-linear LM optimization procedures required to reach a fundamental matrix having a better accuracy than a predefined level is shown in Table 2. For each pair of cameras, we executed 500K RANSAC iterations resulting in 500K hypotheses. Every 1000 RANSAC iterations the best hypothesis is selected and optimized non-linearly. The accuracy of the optimized fundamental matrix, in terms of the symmetric epipolar distance of ground truth points, is recorded. The accuracy values after all non-linear optimization procedures from all camera pairs in the dataset form our samples. For example, in the Kung-Fu dataset we executed $500K \times 300$ RANSAC iterations, performed 500×300 LM optimizations, and collected 150,000 samples. We build the cumulative distribution function (cdf) of the error from all camera pairs. Given the cdf the expected number of samples is extracted. It can be seen that our method quickly converged to sub-pixel accuracy. Fig. 8 shows the ratio between the required number of non-linear optimization procedures in our method and Sinha[28]. The horizontal dashed lines are in ratios of 10, 30 and 100. For accuracy of 0.8 pixel, the median of the ratios between the required number of non-linear optimization procedures is 38, and for accuracy of 1.5 pixel the median of the ratios is 17.

Accuracy We evaluated the best accuracy (minimal error) reached for a given number of RANSAC generated hypotheses. For each pair of cameras in the dataset, we generated 500K RANSAC hypotheses by each method. We sub-

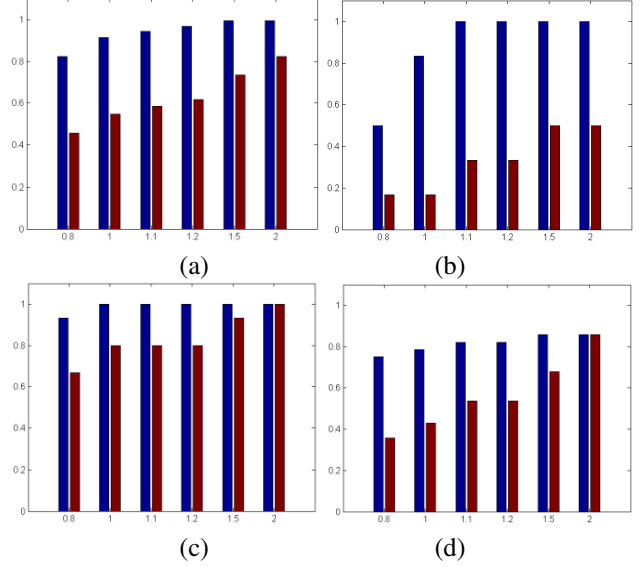


Figure 9. The fraction of camera pairs whose fundamental matrices reached a given symmetric epipolar distance. The accuracy is evaluated over 500K RANSAC iterations. The x-axis is the given accuracy. The y-axis is the fraction of camera pairs that reached this accuracy. The blue bars are our method and the red bars are Sinha’s method. (a) The Kung-Fu dataset. (b) Boxer dataset. (c) Street Dancer. (d) Dancing Girl.

divided the hypotheses into equal sized groups. From each group we selected the best hypothesis (lowest symmetric epipolar distance) with respect to the ground truth points. We then applied non-linear optimization and measured the accuracy of the resulting fundamental matrix. The accuracy is the median over all optimized fundamental matrices. For example, in the Kung-Fu dataset we have 150,000 hypotheses. For evaluation of the highest accuracy reached by 5K hypotheses, we divided them into 30 equal size groups, optimized the best hypothesis from each group and evaluated the median over the symmetric epipolar distances. Table 3 shows the results. It can be seen that for the Kung-Fu dataset, our method requires approximately 2K RANSAC iterations followed with 1 non-linear optimization procedure to reach an accuracy of 0.85. Using our approach, in less than 5K RANSAC iterations all datasets reached sub-pixel accuracy. Table 4 shows the best median accuracy reached by each method. As expected, the synthetic dataset has best accuracy, 0.26, while the worst accuracy, 0.41, was in the Dancing Girl dataset which has many errors in the silhouettes.

We also evaluated the fraction of the number of camera pairs whose fundamental matrices reached a given accuracy using all the samples, after the non-linear phase. The results are shown in Fig. 9. For the Kung-Fu dataset, for 298 out of 300 camera pairs the accuracy reached 1.5, including pairs where the cameras are facing each other. This is discussed

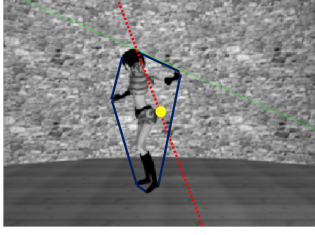


Figure 10. When the epipole is at the center of the image, e.g. when two cameras are facing one another, it may not be possible to find epipolar lines. In this case the epipole is often inside the convex hull. In this example the convex hull is marked in blue, and the yellow point is the epipole. The red line is a ground truth epipolar line. The green line is an hypothesized epipolar line.

in the next subsection. On average, the number of camera pairs where a given accuracy was reached using our method is by a factor of 1.8 higher than the number of cameras with same accuracy using Sinha’s method. The average is calculated over all camera pairs over all datasets.

6.1. Frames Lacking Frontier Points

Frames that lack frontier points are problematic for most tangent based methods. This happens when the epipoles are inside the convex hull of the dynamic objects, a common case when the two cameras face each other. An example is illustrated in Fig. 10. Using our method, even when the pairs of cameras are facing each other, the fundamental matrix can still be recovered. This is possible as the object is moving, and there are often a few frames where the epipole is outside the convex hull. These few frames are enough for the calibration, due to the accuracy of the selected candidates for epipolar lines. For example, it can be seen in Fig. 9 that in the Kung-Fu dataset, for accuracy of 1.5, our method fails for only two camera pairs, whereas Sinha’s method fails on 78 camera pairs.

6.2. Ground-Truth Error vs. Inlier Error

The symmetric epipolar distance [17] is a quality measure for fundamental matrices, and is defined over a set of pairs of corresponding points across two images.

In ordinary computations of the fundamental matrix, when no ground truth data is known, the symmetric epipolar distance is calculated based on *hypothesized* inlier points. Since some inliers are often wrong correspondences, there is a significant difference between the error computed on inlier points and the error computed on ground truth points (when available). As we have access to the ground truth in our datasets, we used the ground truth points to measure the symmetric epipolar distance and evaluate our experiments.

6.3. Implementation Details

Precomputation of Motion Barcodes. The motion bar-

codes were computed for points on the silhouette boundaries which are also on the convex hull boundary, called *candidate points*. 180 angles are sampled every 2° , each angle defines a tangent line to the silhouette through one of the candidate points. This results in 180 candidate lines per frame, and a motion barcode is computed for all these lines. In a video having N frames, each motion barcode is a binary sequence of length N . A barcode matrix is defined for each frame having 180 rows and N columns. Each column represents a frame, and each row represents a tangent line. Each row is the motion barcode of the corresponding candidate line. Given corresponding frames of two cameras, the distance between all possible pairs of candidate lines is computed by multiplying their motion barcode matrices, resulting in an 180×180 affinity matrix of candidate lines.

Given the N pairs of frames of two cameras, we extract for each frame the single pair of candidate lines having the highest barcode correlation. This results in N pairs of possible matching epipolar lines, each having higher barcode correlation.

RANSAC Sampling. The efficiency of fundamental matrix computation can be broken into the initialization cost, the number of hypotheses needed to be generated, the cost of generating a hypothesis and the cost of hypothesis verification. In both methods the cost of the model verification phase is identical as it is indifferent to the model generation. The comparison is therefore the number of RANSAC hypotheses required by each method. In the following we provide a detailed description.

Generating the hypothesized model is as follows. In our method, three matching pairs having high barcode correlation were randomly selected from the pre-computed table of barcode correlations between all pairs of candidate lines. For the Sinha method, as described in [28], two matching hypothesized lines were extracted based on sampling the directions of the tangents in one frame. The third matching pair of lines was computed using the epipole generated by the first two lines, and a tangent to a silhouette in another frame. Given three proposals for corresponding epipolar lines, the fundamental matrix was computed using the method described in [28]. The computation of the third matching pair of lines by the generated epipoles could be applied in our approach as well, requiring selection of only two matching lines instead of three. This could improve the accuracy of the method. On the other hand, it requires additional computations for finding the exact tangents in each RANSAC iteration. We empirically saw that sampling three lines is faster than sampling two lines together with the additional tangent computations.

The cost of each RANSAC iteration depends on (a) lines match generation and (b) the computation of the fundamental matrix from the epipolar line homography and the

epipoles. For the motion barcode method the first part is instantaneous as it involves only index selection, since matching pairs of lines are computed beforehand. For the baseline method each iteration introduces the computation of six tangents, where the computation of the last pair of tangents involves finding the frontier points with respect to the hypothesized epipoles. The second part is the same for all the methods and introduces the major cost of each iteration. We assume that the first part is instantaneous also in the baseline method and consider the cost of each iteration as the cost of the second part.

Computing the motion barcode distance between all pairs of candidate lines adds computation efforts to our method. This computational cost was equivalent to 35 iterations of RANSAC, which we added to the cost of our method.

Ground Truth points. For accurate evaluation of the symmetric epipolar distance we extracted matching frontier points across different views, using the given ground-truth silhouettes and the given ground truth fundamental matrix. For each frame, points whose tangent line is within an angular deviation of 1° of the true epipolar line were extracted. A pair of points was considered frontier if their epipolar distance using the known fundamental matrix is less than 0.01. This results in a cloud of points that might be spread out unevenly. From these points we sampled ground truth points that have a distance of at least 15 pixels from each other, resulting in several dozen point, well spread out, per view.

7. Concluding Remarks

Motion barcodes were introduced as efficient temporal signatures for lines, signatures which are viewpoint invariant for matching epipolar lines. The effectiveness of motion barcodes was demonstrated in camera calibration using candidate epipolar lines. In this case, computing candidate fundamental matrices only from candidate lines that have matching motion barcodes, reduced computational costs by about two orders of magnitude.

Acknowledgment. This research was supported by Google, by Intel ICRI-CI, by DFG, and by the Israel Science Foundation.

References

- [1] Dancer dataset, inria, 4d-repository, <http://4drepository.inrialpes.fr/public/viewgroup/1>, 2010. 5
- [2] Data set, kung-fu girl, <http://www.mpi-inf.mpg.de/departments/irg3/kungfu/>, 2010. 5
- [3] E. Aganj, J.-P. Pons, F. Ségonne, and R. Keriven. Spatio-temporal shape from silhouette using four-dimensional delaunay meshing. In *ICCV'07*, pages 1–8, 2007. 2
- [4] L. Ballan and G. M. Cortelazzo. Multimodal 3d shape recovery from texture, silhouette and shadow information. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 924–930. IEEE, 2006. 5
- [5] G. Ben-Artzi, M. Werman, and S. Peleg. Event retrieval using motion barcodes. In *ICIP'15*, pages 2621–2625, 2015. 1, 2
- [6] E. Boyer. On using silhouettes for camera calibration. In *ACCV'06*, pages 1–10. 2006. 2
- [7] G. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In *CVPR'03*, volume 1, pages 1–77, 2003. 2
- [8] R. Cipolla, K. Astrom, and P. Giblin. Motion from the frontier of curved surfaces. In *ICCV'95*, pages 269–275, 1995. 2
- [9] R. Cipolla and P. Giblin. *Visual motion of curves and surfaces*. Cambridge University Press, 2000. 2
- [10] G. Csürka, C. Zeller, Z. Zhang, and O. D. Faugeras. Characterizing the uncertainty of the fundamental matrix. *CVIU'97*, 68(1):18–36, 1997. 5
- [11] M.-A. Drouin, P.-M. Jodoin, and J. Prémont. Camera-projector matching using an unstructured video stream. In *CVPR'10 Workshop*, pages 33–40. IEEE, 2010. 2
- [12] E. B. Ermis, P. Clarot, P.-M. Jodoin, and V. Saligrama. Activity based matching in distributed camera networks. *IEEE Trans. IP*, 19(10):2595–2613, 2010. 2
- [13] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 2
- [14] K. Forbes, F. Nicolls, G. De Jager, and A. Voigt. Shape-from-silhouette with two mirrors and an uncalibrated camera. In *ECCV'06*, pages 165–178. 2006. 2
- [15] J.-S. Franco and E. Boyer. Exact polyhedral visual hulls. In *BMVC'03*, volume 1, pages 329–338, 2003. 2
- [16] Y. Furukawa, A. Sethi, J. Ponce, and D. Kriegman. Robust structure and motion from outlines of smooth curved surfaces. *IEEE Trans. PAMI*, 28(2):302–315, 2006. 2, 3
- [17] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. 2003. 3, 4, 5, 7
- [18] C. Hernández, F. Schmitt, and R. Cipolla. Silhouette coherence for camera calibration under circular motion. *IEEE Trans. PAMI*, 29(2):343–349, 2007. 2
- [19] Y. Hoshen and S. Peleg. The information in temporal histograms. In *IEEE WACV'15*, pages 924–930, 2015. 2
- [20] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. PAMI*, 16(2):150–162, 1994. 2
- [21] S. Lazebnik, E. Boyer, and J. Ponce. On computing exact visual hulls of solids bounded by smooth surfaces. In *CVPR'01*, volume 1, pages 1–156, 2001. 2
- [22] P. Mendonca, K.-Y. Wong, and R. Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Trans. PAMI*, 23(6):604–616, 2001. 2
- [23] G. Miller and A. Hilton. Exact view-dependent visual-hulls. In *ICPR'06*, pages 107–111, 2006. 2
- [24] J. Porrill and S. Pollard. Curve matching and stereo calibration. *Image and Vision Computing*, 9(1):45–50, 1991. 3

- [25] D. Pundik and Y. Moses. Video synchronization using temporal signals from epipolar lines. In *ECCV'10*, pages 15–28. Springer, 2010. [2](#)
- [26] P. Ramanathan, E. Steinbach, and B. Girod. Silhouette-based multiple-view camera calibration. In *Proc. Vision, Modeling, and Visualization*, pages 3–10, 2000. [2](#)
- [27] C. Schmid and A. Zisserman. The geometry and matching of curves in multiple views. In *ECCV'98*, pages 394–409. 1998. [2](#)
- [28] S. Sinha and M. Pollefeys. Camera network calibration and synchronization from silhouettes in archived video. *IJCV*, 87(3):266–283, 2010. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#)
- [29] J. Starck and A. Hilton. Surface capture for performance-based animation. *Computer Graphics and Applications*, 27(3):21–31, 2007. [5](#)
- [30] A. Stojanovic and M. Unger. A new evaluation criterion for point correspondences in stereo images. In N. Adami, A. Cavallaro, R. Leonardi, and P. Migliorati, editors, *Analysis, Retrieval and Delivery of Multimedia Content*, pages 183–202. Springer, 2013. [5](#)
- [31] K.-Y. Wong and R. Cipolla. Structure and motion from silhouettes. In *ICCV'01*, volume 2, pages 217–222, 2001. [2](#)
- [32] G. Xu and Z. Zhang. *Epipolar geometry in stereo, motion and object recognition: a unified approach*, volume 6. Springer Science & Business Media, 1996. [5](#)
- [33] H. Yamazoe, A. Utsumi, and S. Abe. Multiple camera calibration with bundled optimization using silhouette geometry constraints. In *ICPR'06*, volume 3, pages 960–963. IEEE, 2006. [2](#)
- [34] H. Zhang and K.-Y. Wong. Self-calibration of turntable sequences from silhouettes. *IEEE Trans. PAMI*, 31(1):5–14, 2009. [2](#)